

# Gaussian Processes for Blazar Variability Studies

Vassilis Karamanavis

Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-53121 Bonn, Germany;  
vkaramanavis@mpifr.de

Academic Editor: name

Received: date; Accepted: date; Published: date

**Abstract:** This article shortly introduces Gaussian processes as a new approach for modelling time series in the field of blazar physics. In the second part of the paper, recent results from an application of GP modelling to the multi-wavelength light curves of the blazar PKS 1502+106 are discussed.

**Keywords:** active galaxies; blazars; jets; radio continuum; gamma-rays; time-series analysis; Bayesian modelling; Gaussian processes

## 1. Introduction

Detailed investigations of the emission from blazars have offered, and continue to offer, important insights into this extreme population of active galactic nuclei (AGN). Blazar variability is an indispensable tool because it offers information on scales even smaller than those probed with imaging techniques. Furthermore, given the predictions of various blazar variability models [e.g., 1–4], we are able to assess their validity taking advantage of the fingerprints of each model in the time and frequency domains, namely on light curves and broadband spectra.

Numerous monitoring efforts have been in place since the early days of the discovery of AGN and blazars, however most, if not all, lacked the high time resolution required for detailed studies. The tables have turned with the launch of the *Fermi*  $\gamma$ -ray space telescope and complementary efforts for multi-frequency blazar monitoring. One such example in the radio domain is the F-GAMMA (*Fermi*-GST AGN Multi-frequency Monitoring Alliance) program [5,6]. The F-GAMMA operated between 2007 and 2015 and served as a prime provider of single-dish radio monitoring complementary to *Fermi* [7].

The goal of variability studies is the extraction of parameters such as the variability amplitudes, time scales, and others, from observations. The problem of modelling the observed variability in the radio and other bands, has been addressed using numerous approaches. For modelling light curves and the flares seen in them, Gaussian, exponential, and various other fitting methods have been used (e.g. [8–10]; this is by no means a complete list). Depending on each specific case, they work better or worse but the common characteristic among all the above is that a functional form of the underlying process is always assumed. In other words, a parametric approach is being employed. However, this might not be appropriate and we might have to admit that at times we have weak prior knowledge as to what is the mathematical form of such a process.

An elegant way to overcome this state of limited knowledge is by means of Bayesian inference. Specifically, this paper is concerned with the application of Gaussian processes in the field of blazar variability studies and modelling of time series. In the next section, a brief introduction to Gaussian processes is given. Then, an application to the light curves of the blazar PKS 1502+106 is discussed, aiming to quantify an outburst seen from 2008 to 2010 and extract light curve parameters like the flare amplitude and putative delays between its peak at different bands. First results along with an extended discussion can be found in [11].

## 2. Non-parametric models and Gaussian processes

The problem at hand is as follows: we want to extract (or learn, in the context of Bayesian machine learning) the properties of a system from a given set of observations. This can be a function

$f()$  that describes the system well enough and predicts its future evolution. There are essentially two ways of achieving this goal. The first involves the selection of a class of functions, e.g. linear, quadratic, Gaussian etc. with (free) parameters, the values of which are obtained through minimizing the residuals between  $f()$  and our observations. Such a selection can result from strong prior knowledge regarding the system. One can increase the flexibility of the model by increasing the number of parameters at the expense, though, of potential overfitting; i.e. describing random small scale variations of the data instead of the general trend.

Still, cases in which we have little or no prior knowledge regarding the underlying process, and consequently the appropriate model to use, exist in abundance. Non-parametric models offer an interesting alternative to the first approach. By comparison, non-parametric models “translate” our prior knowledge about a system, no matter how generic this might be, e.g. smooth and/or continuous changes, into probability distributions of functions that are more likely to describe the system. The task of modelling then reduces to fine-tuning these prior distributions so that they include functions that best describe our observations; i.e. the posterior distributions. Since no explicit use of free parameters is made, these models are referred to as non-parametric. Non-parametric models, however, use so-called hyperparameters which instead of determining the properties of functions, as in the first approach, govern the properties of the distribution. These still need to be inferred and Bayesian model selection provides a robust framework for doing so.

Gaussian processes (GPs) are the generalization of uni- or multivariate Gaussian distributions of variables into the space of functions. GPs offer a very flexible framework for modelling unknown functions [12–14]. They are widely used in machine learning applications and a new field of applications in astronomy has recently opened [e.g. 15–18]. The interested reader is referred to [19] for additional details.

### 2.1. Gaussian processes for regression

Given a set of observations, e.g. a typical light curve, with the form  $D = (x_i, y_i)$ , where  $y_i$  represent recorded flux densities at times  $x_i$ , we would like to find all functions drawn from the prior distribution which pass through all observations. The latter distribution would be the posterior distribution and its mean value describes the data best (see Figure 1).

What generates the prior distribution is a covariance kernel or function. This governs the similarity (covariance) between any two function values, at  $x_i$  and  $x_j$  say. It is chosen based on our prior knowledge of the system. There exists a multitude of kernels which can model unknown functions based on their characteristics, e.g. periodic, linear etc. [e.g. 20]. It is noteworthy that sums and products of kernels produce valid kernels too. For the purposes of this paper, the squared-exponential (SE) kernel is chosen. This is the simplest kernel and the only underlying assumption is that of a smooth process (i.e. the functions it produces have derivatives of all orders everywhere in their domain). The SE kernel is defined as

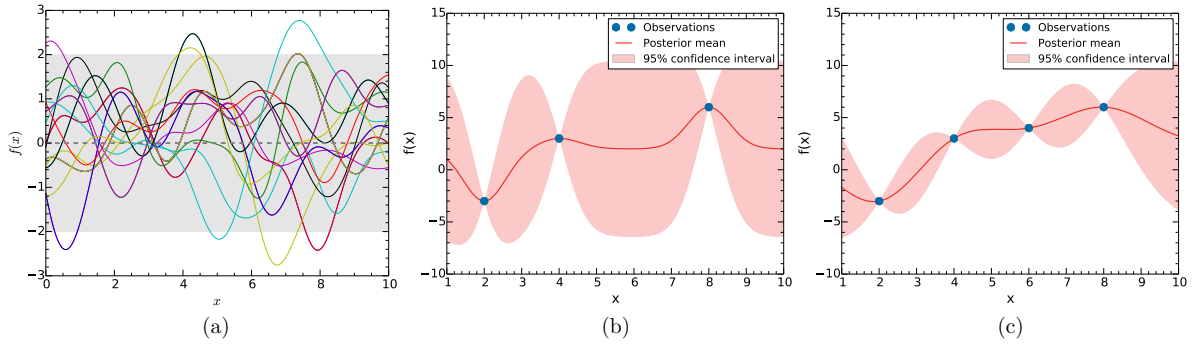
$$k(x_i, x_j) = \sigma_f^2 \exp \left[ \frac{-(x_i - x_j)^2}{2l^2} \right], \quad (1)$$

with  $l$  the characteristic length scale, i.e. the distance in the  $x$  dimension after which the function changes significantly and  $\sigma_f^2$ , the variance, i.e. the mean distance of the function from its mean (serves only as a scaling factor). Here,  $l$  and  $\sigma_f^2$  are the hyperparameters.

Uncertainty is accommodated by simply adding a term describing it, to the noiseless SE covariance kernel of Eq. 1, which then becomes

$$\mathbf{k}_n(x_i, x_j) = \mathbf{k}(x_i, x_j) + \sigma^2 \mathbf{I}, \quad (2)$$

with  $\mathbf{I}$  the identity matrix [e.g. 13].



**Figure 1.** Gaussian process prior and posterior distributions: (a) Twenty randomly selected samples from the prior distribution, with zero mean and  $\sigma = 1$ , using a squared exponential covariance kernel. (b) Regression results after training the GP with three error-free data points. (c) Same as in panel (b), however note the improved result with the addition of one more data point.

## 2.2. Training the Gaussian process

Bayesian model selection provides the way for optimizing the hyperparameters, thereby refining the posterior distribution and enabling us to perform inference. We essentially update our prior knowledge in the light of a data set. This is done by means of maximizing the log marginal likelihood [12], as outlined below.

For the SE kernel, the vector of hyperparameters is  $\theta = \{l, \sigma^2\}$  and the probability (or evidence) of the training data  $\mathbf{y}$ , given the hyperparameters vector  $\theta$ , is  $p(\mathbf{y} | \mathbf{x}, \theta)$ . The log marginal likelihood is given by

$$\mathcal{L} = \log p(\mathbf{y} | \mathbf{x}, \theta) = -\frac{1}{2} \mathbf{y}^T \mathbf{k}_n^{-1} \mathbf{y} - \frac{1}{2} \log |\mathbf{k}_n| - \frac{n}{2} \log 2\pi. \quad (3)$$

More generally, in the case of a hyperparameter vector  $\theta = \{\theta_j | j = 1, \dots, n\}$ , the derivatives of the log marginal likelihood with respect to each  $\theta_j$  are

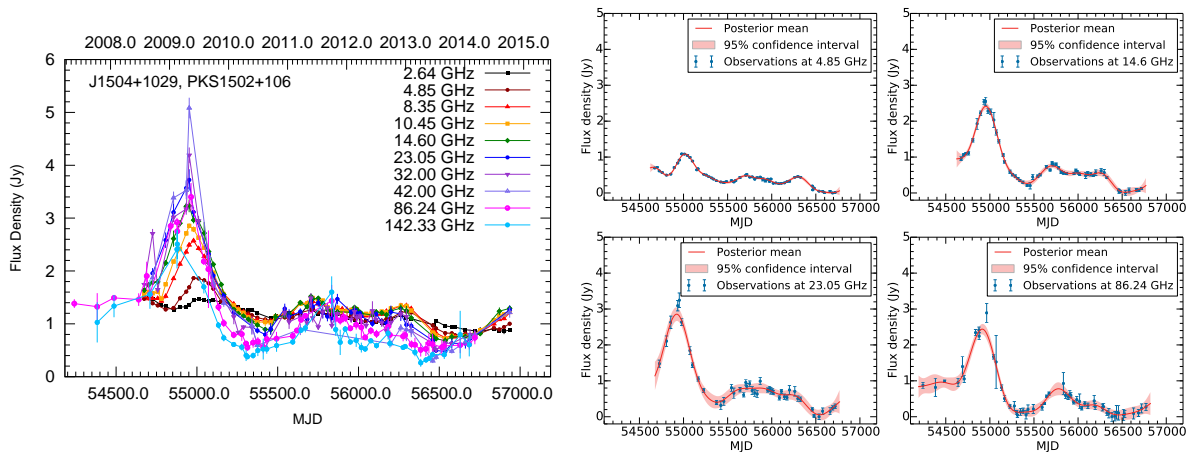
$$\frac{\partial \mathcal{L}}{\partial \theta_j} = \frac{1}{2} \mathbf{y}^T \frac{\partial \mathbf{k}_n}{\partial \theta_j} \mathbf{k}_n^{-1} \mathbf{y} - \frac{1}{2} \text{Trace} \left( \mathbf{k}_n^{-1} \frac{\partial \mathbf{k}_n}{\partial \theta_j} \right). \quad (4)$$

By means of numerical gradient optimization algorithms one can maximize the log marginal likelihood using Eq. 4 and obtain the best set of hyperparameters (for an outline of the algorithm see [19]).

## 3. Application to the light curves of the blazar PKS 1502+106

PKS 1502+106 (S3 1502+10, OR 103) is a powerful flat-spectrum radio quasar (FSRQ), driven by a supermassive black hole of about  $10^9 M_\odot$ , at redshift  $z = 1.8385$  (see [21] and references therein). The source has been detected by *Fermi* in 2008 showing an intense  $\gamma$ -ray outburst at MeV–GeV energies [22]. The flare was monitored by F-GAMMA and other programs in the radio band. Radio data from the F-GAMMA program<sup>1</sup> were employed, including observations with the Effelsberg 100-m (EB) and the IRAM 30-m (PV) telescopes in ten frequency bands from 2.64 to 142.33 GHz. Monthly observations from the EB and PV telescopes were performed quasi-simultaneously, thus ensuring maximum spectral coherency. A detailed description of the observational setup and data reduction is provided elsewhere [7,10,23,24].

<sup>1</sup> [www.mpifr-bonn.mpg.de/div/vlbi/fgamma/fgamma.html](http://www.mpifr-bonn.mpg.de/div/vlbi/fgamma/fgamma.html)



**Figure 2.** Multi-wavelength light curves of PKS 1502+106. **(Left)** F-GAMMA light curves between 2.64 and 142.33 GHz. **(Right)** GP regression curves for the data at 4.85, 14.6, 23.05, and 86.24 GHz. For the results at all frequencies see [18].

Considering the full length of available multi-frequency light curves, GP regression was applied to the data at each frequency separately. A variant of the software by [25] was used, which was adapted to our specific modelling needs. A suite of machine learning applications for Python, including GP regression, is available online<sup>2</sup>. To obtain the best unbiased result, the length scale parameter  $l$  was randomly initialized 100 times. Then, the posterior mean is returned with a 95% confidence interval for the flux density, along with the set of hyperparameters which maximize the log marginal likelihood. The relevant light curve parameters extracted are the peak amplitude ( $S_m$ ), the time of the peak, and the cross-band delay. Furthermore, the flare rise and decay times are calculated between the peak and the two flux density minima which immediately precede and follow  $S_m$  [18].

## 4. Results and discussion

### 4.1. Quantifying the broadband outburst of PKS 1502+106

PKS 1502+106 has shown an isolated outburst visible across the electromagnetic spectrum. The results of the GP regression are visible in the right panels of Fig. 2. The amplitudes at each of our radio frequencies first increase, up to about 43 GHz where they peak, and then drop at higher  $\nu$  (Fig. 4 of Ref. [18]). The flare peaks with increasing time delay as we follow it towards lower frequencies (Fig. 5 of Ref. [18]). The dependence of amplitudes and delays on  $\nu$  can be well approximated by power laws, a telltale sign of shock evolution within the source's parsec-scale jet. These results are in agreement with the study of the source using mm-wave very-long-baseline interferometry (mm-VLBI; see [21]).

The observed time lags between radio frequencies can be attributed to synchrotron opacity effects. In this context, the flare maxima appear first at higher frequencies and gradually progress to lower frequencies [e.g. 2]. Timing the flare maximum at each frequency with respect to the data at 142.33 GHz yields the following delays. The longest delay of  $(205 \pm 2.0)$  days is seen between 142.33 and 2.64 GHz. Decreasing lags are clearly visible towards higher frequencies with a time lag of  $(136.9 \pm 8.1)$  d at 4.85 GHz and  $(96.2 \pm 3.8)$  d at 8.35 GHz. At higher frequencies, the lags are  $(85.8 \pm 5.2)$  d at 10.45 GHz,  $(67.2 \pm 0.9)$  d at 14.6,  $(39.4 \pm 5.6)$  d at 23.03, and  $(39.8 \pm 11.2)$  d at 32 GHz. At the two highest frequencies of 43.05 GHz and 86.24 GHz, the lags are  $(40.8 \pm 3.1)$  d and  $(24.8 \pm 0.2)$  d, respectively (Table 3 of Ref. [18]).

<sup>2</sup> [www.scikit-learn.org](http://www.scikit-learn.org)

The aforementioned lags allow the calculation of the synchrotron opacity profile of PKS 1502+106. Considering that at a given frequency the core represents a feature whose optical depth is close to unity, the standard relativistic jet paradigm predicts that the apparent position of this unit-opacity surface depends on observing frequency (see [18] and references therein). The importance of this “core-shift” effect lies in the fact that it gives critical insights into the physical processes of ultracompact jets, like the de-projected distance between the core (at each frequency) and the jet base, as well as the strength of the magnetic field along the flow [e.g., 26,27]. Using a “time-lag core shift” method, which proves to be a good alternative to VLBI core-shifts [e.g. 9,28], the positions of the nine unit-opacity surfaces (or photospheres) with respect to the jet base were inferred. These are:  $(10.2 \pm 1.2)$  pc for the 2.64 GHz core,  $(7.1 \pm 1.0)$  pc at 4.85,  $(5.3 \pm 0.8)$  pc at 8.35,  $(4.8 \pm 0.8)$  pc at 10.45,  $(4.0 \pm 0.7)$  pc at 14.60,  $(2.6 \pm 0.6)$  pc at 23.05,  $(2.9 \pm 1.0)$  pc at 32.00,  $(3.4 \pm 0.8)$  pc at 43.05, and  $(4.0 \pm 1.1)$  pc at 86.24 GHz (see Table 4 of Ref. [18] for the values and their calculation). Finally, we obtain the equipartition magnetic field along the jet axis, with values increasing from 14 mG at the 2.64 GHz core up to 176 mG at 86.24 GHz.

#### 4.2. Where do the $\gamma$ rays come from?

The location of the  $\gamma$ -ray emitting region is decisively constrained by combining the aforementioned findings with those of [23] and [21]. Specifically, in [23] the authors calculate the relative distance between the 86 GHz core and the  $\gamma$ -ray site to be about 2.1 pc, while in [21], using high-resolution mm-VLBI imaging, the knot most likely connected with the flare in question is identified (knot C3) and its kinematical properties are deduced ( $\beta_{\text{app}} \sim 7c$ , jet viewing angle of  $2.6^\circ$ ).

Having calculated the absolute distance between the jet base and the 86.24 GHz core ( $4.0 \pm 1.1$ ) pc, and with the relative separation between the same core and the  $\gamma$ -ray site (2.1 pc), the high-energy emission originates at  $(1.9 \pm 1.1)$  pc from the jet base. This places it far from the bulk of the broad-line material of the source ( $R_{\text{BLR}} \approx 0.1$  pc) making external Compton, on a field other than that of the BLR, a very relevant emission mechanism for PKS 1502+106’s  $\gamma$  ray emission, with implications on the origin of the seed photon field (see [18] and references therein).

**Acknowledgments:** V.K. thanks L. Fuhrmann, E. Angelakis, T. P. Krichbaum, I. Myserlis, I. Nestoras, J. A. Zensus, and H. Ungerechts for interesting discussions and useful comments; also W. Max-Moerbeck and B. Boccardi for carefully reading the paper. Partially based on observations with the Effelsberg 100-m and the IRAM 30-m telescopes.

#### Bibliography

1. Marscher, A.P.; Gear, W.K. Models for high-frequency radio outbursts in extragalactic sources, with application to the early 1983 millimeter-to-infrared flare of 3C 273. *ApJ* **1985**, *298*, 114–127.
2. Valtaoja, E.; Terasranta, H.; Urpo, S.; Nesterov, N.S.; Lainela, M.; Valtonen, M. Five Years Monitoring of Extragalactic Radio Sources - Part Three - Generalized Shock Models and the Dependence of Variability on Frequency. *A&A* **1992**, *254*, 71.
3. Spada, M.; Ghisellini, G.; Lazzati, D.; Celotti, A. Internal shocks in the jets of radio-loud quasars. *MNRAS* **2001**, *325*, 1559–1570, [[astro-ph/0103424](#)].
4. Camenzind, M.; Krockenberger, M. The lighthouse effect of relativistic jets in blazars - A geometric origin of intraday variability. *A&A* **1992**, *255*, 59–62.
5. Fuhrmann, L.; Zensus, J.A.; Krichbaum, T.P.; Angelakis, E.; Readhead, A.C.S. Simultaneous Radio to (Sub-) mm-Monitoring of Variability and Spectral Shape Evolution of potential GLAST Blazars. The First GLAST Symposium; Ritz, S.; Michelson, P.; Meegan, C.A., Eds., 2007, Vol. 921, *American Institute of Physics Conference Series*, pp. 249–251, [[0704.3944](#)].
6. Angelakis, E.; Fuhrmann, L.; Nestoras, I.; Zensus, J.A.; Marchili, N.; Pavlidou, V.; Krichbaum, T.P. The F-GAMMA program: multi-wavelength AGN studies in the Fermi-GST era. *ArXiv e-prints* **2010**, [[arXiv:astro-ph.CO/1006.5610](#)].



7. Fuhrmann, L.; Angelakis, E.; Zensus, J.A.; Nestoras, I.; Marchili, N.; Pavlidou, V.; Karamanavis, V.; Ungerechts, H.; Krichbaum, T.P.; Larsson, S.; Lee, S.S.; Max-Moerbeck, W.; Myserlis, I.; Pearson, T.J.; Readhead, A.C.S.; Richards, J.L.; Sievers, A.; Sohn, B.W. The F-GAMMA program: Multi-frequency study of Active Galactic Nuclei in the Fermi era. Program description and the first 2.5 years of monitoring. *ArXiv e-prints* **2016**, [[arXiv:astro-ph.HE/1608.02580](https://arxiv.org/abs/1608.02580)].
8. Valtaoja, E.; Lähteenmäki, A.; Teräsanta, H.; Lainela, M. Total Flux Density Variations in Extragalactic Radio Sources. I. Decomposition of Variations into Exponential Flares. *ApJS* **1999**, *120*, 95–99.
9. Kudryavtseva, N.A.; Gabuzda, D.C.; Aller, M.F.; Aller, H.D. A new method for estimating frequency-dependent core shifts in active galactic nucleus jets. *MNRAS* **2011**, *415*, 1631–1637, [[arXiv:astro-ph.HE/1106.0069](https://arxiv.org/abs/1106.0069)].
10. Angelakis, E.; Fuhrmann, L.; Marchili, N.; Foschini, L.; Myserlis, I.; Karamanavis, V.; Komossa, S.; Blinov, D.; Krichbaum, T.P.; Sievers, A.; Ungerechts, H.; Zensus, J.A. Radio jet emission from GeV-emitting narrow-line Seyfert 1 galaxies. *A&A* **2015**, *575*, A55, [[arXiv:astro-ph.HE/1501.02158](https://arxiv.org/abs/1501.02158)].
11. Karamanavis, V. Zooming into  $\gamma$ -ray loud galactic nuclei: broadband emission and structure dynamics of the blazar PKS 1502+106 and the narrow-line Seyfert 1 1H 0323+342. PhD thesis, Universität zu Köln, 2015.
12. Williams, C.K.I.; Rasmussen, C.E. Gaussian Processes for Regression. *Advances in Neural Information Processing Systems 8*. MIT press, 1996, pp. 514–520.
13. Roberts, S.; Osborne, M.; Ebdn, M.; Reece, S.; Gibson, N.; Aigrain, S. Gaussian processes for time-series modelling. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **2012**, *371*, [<http://rsta.royalsocietypublishing.org/content/371/1984/20110550.full.pdf>].
14. Ghahramani, Z. Probabilistic machine learning and artificial intelligence. *Nature* **2015**, *521*, 452–459.
15. Gibson, N.P.; Aigrain, S.; Roberts, S.; Evans, T.M.; Osborne, M.; Pont, F. A Gaussian process framework for modelling instrumental systematics: application to transmission spectroscopy. *MNRAS* **2012**, *419*, 2683–2694, [[arXiv:astro-ph.EP/1109.3251](https://arxiv.org/abs/1109.3251)].
16. Aigrain, S.; Pont, F.; Zucker, S. A simple method to estimate radial velocity variations due to stellar activity using photometry. *MNRAS* **2012**, *419*, 3147–3158, [[arXiv:astro-ph.SR/1110.1034](https://arxiv.org/abs/1110.1034)].
17. Rajpaul, V.; Aigrain, S.; Osborne, M.A.; Reece, S.; Roberts, S. A Gaussian process framework for modelling stellar activity signals in radial velocity data. *MNRAS* **2015**, *452*, 2269–2291, [[arXiv:astro-ph.EP/1506.07304](https://arxiv.org/abs/1506.07304)].
18. Karamanavis, V.; Fuhrmann, L.; Angelakis, E.; Nestoras, I.; Myserlis, I.; Krichbaum, T.P.; Zensus, J.A.; Ungerechts, H.; Sievers, A.; Gurwell, M.A. What can the 2008/10 broadband flare of PKS 1502+106 tell us?. Nuclear opacity, magnetic fields, and the location of  $\gamma$  rays. *A&A* **2016**, *590*, A48, [[arXiv:astro-ph.HE/1603.04220](https://arxiv.org/abs/1603.04220)].
19. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*; The MIT Press, 2005.
20. Duvenaud, D. Automatic Model Construction with Gaussian Processes. PhD thesis, Computational and Biological Learning Laboratory, University of Cambridge, 2014.
21. Karamanavis, V.; Fuhrmann, L.; Krichbaum, T.P.; Angelakis, E.; Hodgson, J.; Nestoras, I.; Myserlis, I.; Zensus, J.A.; Sievers, A.; Ciprini, S. PKS 1502+106: A high-redshift Fermi blazar at extreme angular resolution. Structural dynamics with VLBI imaging up to 86 GHz. *A&A* **2016**, *586*, A60, [[arXiv:astro-ph.HE/1511.01085](https://arxiv.org/abs/1511.01085)].
22. Abdo, A.A.; Ackermann, M.; Ajello, M.; Atwood, W.B.; Axelsson, M.; Baldini, L.; Ballet, J.; Barbiellini, G.; Bastieri, D.; Baughman, B.M.; et al. PKS 1502+106: A New and Distant Gamma-ray Blazar in Outburst Discovered by the Fermi Large Area Telescope. *ApJ* **2010**, *710*, 810–827, [[arXiv:astro-ph.HE/0912.4029](https://arxiv.org/abs/0912.4029)].
23. Fuhrmann, L.; Larsson, S.; Chiang, J.; Angelakis, E.; Zensus, J.A.; Nestoras, I.; Krichbaum, T.P.; Ungerechts, H.; Sievers, A.; Pavlidou, V.; Readhead, A.C.S.; Max-Moerbeck, W.; Pearson, T.J. Detection of significant cm to sub-mm band radio and  $\gamma$ -ray correlated variability in Fermi bright blazars. *MNRAS* **2014**, *441*, 1899–1909, [[arXiv:astro-ph.HE/1403.4170](https://arxiv.org/abs/1403.4170)].
24. Nestoras, I.; et al.; . F-GAMMA program: Multi-frequency study of Active Galactic Nuclei in the Fermi-GST era III. The first 5 years of IRAM 30-m monitoring. *A&A* *submitted*.

25. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.
26. Lobanov, A.P. Ultracompact jets in active galactic nuclei. *A&A* **1998**, *330*, 79–89.
27. Hirokani, K. Kinetic Luminosity and Composition of Active Galactic Nuclei Jets. *ApJ* **2005**, *619*, 73–85.
28. Kutkin, A.M.; Sokolovsky, K.V.; Lisakov, M.M.; Kovalev, Y.Y.; Savolainen, T.; Voytsik, P.A.; Lobanov, A.P.; Aller, H.D.; Aller, M.F.; Lahteenmaki, A.; Tornikoski, M.; Volvach, A.E.; Volvach, L.N. The core shift effect in the blazar 3C 454.3. *MNRAS* **2014**, *437*, 3396–3404, [[arXiv:astro-ph.HE/1307.4100](https://arxiv.org/abs/1307.4100)].

© 2016 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).